

"Laute - Wörter - Sätze - Text": EDV-gestützte Herstellung eines Lehrbuches für das Alt- und Mittelenglische

Kemmler, Fritz

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Kemmler, F. (1995). "Laute - Wörter - Sätze - Text": EDV-gestützte Herstellung eines Lehrbuches für das Alt- und Mittelenglische. *Historical Social Research*, 20(1), 106-109. <https://doi.org/10.12759/hsr.20.1995.1.106-109>

Nutzungsbedingungen:

Dieser Text wird unter einer CC BY Lizenz (Namensnennung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:
<https://creativecommons.org/licenses/by/4.0/deed.de>

Terms of use:

This document is made available under a CC BY Licence (Attribution). For more Information see:
<https://creativecommons.org/licenses/by/4.0>

HUMANITIES COMPUTING

»Laute - Wörter - Sätze - Text«: EDV-gestützte Herstellung eines Lehrbuches für das Alt- und Mittelenglische

*Fritz Kemmler (Tübingen)**

Im vorgestellten Projekt geht es um den Einsatz und den Nutzen der EDV bei den vielfältigen Arbeiten, die mit der Herstellung eines Lehrbuches für das Alt- und Mittelenglische verbunden sind. Der Begriff »Herstellung« bezieht sich dabei sowohl auf die Inhalte wie auch auf die Techniken und Probleme des elektronischen Setzens.

Das im Herbst 1994 erscheinende Lehrbuch wird folgende Gliederung haben:

- eine Grammatik mit Abschnitten zur Lautlehre, Formenlehre und Syntax für das Alt- und Mittelenglische
- zehn Texte und Textausschnitte aus dem Bereich der altenglischen Literatur
- vierzehn Texte und Textausschnitte aus dem Bereich der mittelenglischen Literatur
- Literaturverzeichnis
- Glossare für die alt- und mittelenglischen Texte.

Jeder dieser Teile bringt ganz spezifische satztechnische Probleme mit sich:

- Die Lektüretex te z. B. sollen einen Zeilenzähler erhalten, der gleichzeitig Bestandteil der Referenz in den Glossaren ist
- Sprachliche und sprachgeschichtliche Erläuterungen zu einzelnen Formen des Textes sollen als Einschaltung (Wechsel der Schriftgröße) vor dem jeweiligen Text stehen - mit Verweisen auf die Zeilen, in der sie im Text zu finden sind.
- Inhaltliche und sprachliche Erläuterungen zu schwierigen Textstellen sollen in der Form von Fußnoten bei den Texten stehen, jedoch soll die Fußnotenverweisziffer weder im Text noch bei der Fußnote erscheinen.
- Im Glossar für die altenglischen Texte sollen sämtliche Formen aufgeführt sein; das Glossar für die mittelenglischen Texte ist dagegen ein Auswahlglossar.

* Protokoll des 60. Kolloquiums über die Anwendung der EDV in den Geisteswissenschaften an der Universität Tübingen am 12. Februar 1994.

- Im Bereich der Formenlehre im Grammatik-Teil sind mit erheblichem Arbeitsaufwand verbundene Konjugations- und Deklinationsparadigmata zu erstellen usw.

Die für diese Arbeiten nach und nach erstellten TUSTEP-Programme sind zum einen im Bereich #SATZ angesiedelt, zum anderen im Bereich der eigentlich sprachhistorischen Fragestellungen und Probleme, also im Bereich der Inhalte.

I.

Für die Erfassung und Bearbeitung des Literaturverzeichnisses wurde eine Prozedur in Form eines Benutzermakros zusammengestellt. Dieses dient einerseits der Vermeidung formaler Fehler bei der Eingabe der Daten, andererseits der Plausibilitätsprüfung der Daten; es erlaubt weiterhin eine mit wenig Aufwand verbundene alphabetische Sortierung der Einträge und die Ergänzung der für das Satzprogramm notwendigen Steueranweisungen.

Die tabellarisch angeordneten Konjugations- und Deklinationsparadigmata werden mit dem Satz-Programm für den Druck aufbereitet. Da die in einer solchen Tabelle stehenden Formen unterschiedlich lang sind, ist es nicht ganz einfach, ohne die Hilfe eines formatierenden Editor-Bildschirmes die zur Herstellung einer solchen Tabelle notwendigen Merk- und Positionierungsanweisungen zu definieren. Um in diesem Bereich Doppelbelichtungen zu vermeiden und um die Tabellen auch optisch einigermaßen ansprechend gestalten zu können, werden die Leistungen der Anweisung »&!e« in der Protokoll-Datei eines Satzlaufes systematisch ausgewertet. Zur Umrechnung der in der Protokoll-Datei in Bildlinien angegebenen horizontalen Position in Didot-Punkt (und darauf fußen ja die Merk- und Positionierungsanweisungen) dient ein Benutzermakro, das außerdem im Dialog Informationen zur verwendeten Schriftgröße berücksichtigen kann.

Die im Buch abgedruckten Texte enthalten Erläuterungen zu schwierigen Textstellen in der Form von Fußnoten (jedoch ohne Fußnotenverweisziffer im Text und bei der Fußnote). Ein Programm ermöglicht es, die Fußnoten maschinell mit eindeutigen Verweisen zu versehen, so daß der Benutzer die in den »Fußnoten« stehenden Informationen der spezifischen Stelle im Text eindeutig zuordnen kann. Der Ansatzpunkt dabei ist der in den einzelnen Texten fortlaufende Zeilenzähler; die jeweilige Zeilennummer wird von der Maschine hinter der Fußnotenverweisziffer eingetragen. Bei mehr als 500 solcher »Fußnoten« in den Texten kann es so bei den Verweisen zu keinen Eingabefehlern (Zahlen-drehern) kommen.

II.

Bei der Herstellung der Glossare waren zu folgenden Fragen Lösungen zu finden:

- Wie können die Kennzeichnungen für vokalische Länge (TUSTEP-Codierung »%-«) bei der Sortierung berücksichtigt werden, wenn Länge nach Kürze stehen soll?
- Wie sind die **Sonderzeichen** (z. B. »aesh«, »thorn/eth«) zu behandeln?
- Wie kann etwa bei den Pronominalformen und bei den Konjugationsformen eine Vorsortierung durch die Maschine erreicht werden? Um nur ein Beispiel anzuführen: die im Präsens konjugierten Formen des Verbes »gehen« beginnen im Mittenglischen mit dem Buchstaben »g-«; im Präteritum konjugierte Formen erscheinen in den Texten aber mit »3-, w-« und »y-«. Bei den altenglischen Texten ergeben sich vergleichbare Sortierprobleme.
- Wie kann **man erreichen, daß** etwa **die Belege** »aldorman« und »aldormon« aus den altenglischen Texten den gleichen Sortierwert erhalten, in der Arbeitsdatei also unter »ea-« einsortiert sind?
- Ist es möglich, bei der Herstellung der Arbeitsdatei für die Glossare bei bestimmten Register-Einträgen (etwa bei den problematischen »Wörtchen« *a, i, o, ar, er, or*) auch noch genügend vom syntaktisch-semanticen Kontext zu berücksichtigen, so daß man bei der Bearbeitung der Glossare nicht ständig von der Belegdatei in die Textdatei wechseln muß?

Für diese arbeitstechnischen Probleme wurden TUSTEP-Programme entwickelt, die befriedigende Ergebnisse liefern.

III.

Auch bei philologischen Fragestellungen im engeren Sinne kann die Maschine Hilfe bieten:

Im Bereich der Syntax (Stichwort »Sätze«) werden mit Hilfe eines Programmes bestimmte syntaktische Strukturen in den Texten abgefragt, isoliert und auf verschiedene Zielformate ausgegeben, die z. B. nur Relativsätze, Temporalsätze usw. enthalten sollen. Das Datenformat der dieser Auswahl zugrunde liegenden Quelldatei sieht folgendermaßen aus:

<SA406#32> Ne þearf ðær þonne ænig anforht wasan, þe him ær on brēostum bereð bæacna sēlest; ac ðurh ðā rōde sceal rīce gesēcan of eorðwege æghwylc sǽwl, sēo þe mid Wealdende wunian þenceð. <E-R> <SE>

Die Informationen in »< >« vor dem Satz geben Auskunft über die laufende Nummer der syntaktischen Einheit sowie über ihre Länge in Wörtern. Der mitlaufende Wortzähler ermöglicht u. a. auch stilistische Untersuchungen.

Auf dem Gebiet der sprachlichen Merkmale (Stichwort »Laute«), die zur Bestimmung der regionalen Herkunft besonders eines mittenglischen Textes dienlich sein können, werden zwei Programme eingesetzt, die solche Merkmale im großen und ganzen recht zuverlässig erkennen und für die endgültige philologische Interpretation und Bewertung - und diese Arbeit wird die Maschine wohl nie leisten können - mit einem Kennzeichen versehen (machine-tagging). So ist es mit Hilfe der Maschine durchaus möglich, recht schnell eine erste

Orientierung über die regionale Herkunft eines Textes zu bekommen. Das Programm zur Filterung und Markierung regionaler sprachlicher Merkmale kann romanisch/lateinische und skandinavische Lehnwörter in mitttelenglischen Texten vorläufig markieren, wie das folgende Beispiel zeigt, dem ein kentischer Text aus der 1. Hälfte des 14. Jahrhunderts zugrunde hegt.

Markierungen durch #KOPIERE:

Süden?

@ ariseþ [1] 66

Kent?

Süden?

@ arizeþ [1] 51

Die Zahlen in eckigen Klammern geben die Häufigkeit des Wortes an, die dahinter stehenden Zahlen die Zeilennummer des zugrunde liegenden Textes.

Ergebnisse der Bearbeitung:

1001 Wörter im Text

372 Registereinträge

157 Registereinträge markiert

Besondere Merkmale des Textes:

a) Graphien und Wortschatz (40 Markierungen):

23 mal in- u. auslautendes »yogh«

15 mal Markierung für roman. Lehnwort

2 mal Markierung für skand. Lehnwort

b) Leitformen für die Dialekte (116 Markierungen):

47 für den Süden

3 für das westliche Mittelland und den Südwesten

54 für Kent

5 für Kent und den Südwesten

5 für den Südosten

1 für das östliche Mittelland

1 für das westliche Mittelland

0 für den Norden

Südhumbrischer Text